CS 395T: Sublinear Algorithms Fall 2016

Lecture 12 — October 4, 2016

Prof. Eric Price *Scribe: Akshay D Kamath, Surbhi Goel*

1 Lower Bounds for p^{th} Moment ($p > 2$)

We will prove lower bounds for p^{th} Moment ($p > 2$) using information theory. Suppose we have an algorithm Alg and a distribution on $A \in \mathbb{R}^{m\times n}$ such that given $x \in \mathbb{R}^n$, running Alg on (Ax) outputs v that satisfies $v = (1 \pm \epsilon)||x||_p^p$ w.p. $1 - \delta$.

Consider the following setup,

- Alice has bit $b \in \{0, 1\}$ drawn uniformly at random. She chooses $x^*, w \in \mathbb{R}^n$ such that $w \sim N(0, \sigma^2 I_n)$ and

$$x^* = \begin{cases} 0 & \text{if } b = 0\,, \\ e_i & \text{w.p. } 1/n \ \forall i \in \{1, \dots, n\} \text{ if } b = 1\,. \end{cases}$$

  She sends $A(x^* + w)$ to Bob.

- Bob runs Alg on $A(x^* + w)$ to estimate $||x^* + w||_p^p$ and outputs $\hat{b} = Alg(A(x^* + w)) > 1.5$.

We can easily see that,

$$||x^* + w||_p^p \approx \begin{cases} n\sigma^p & \text{if } b = 0\,, \\ 1 + n\sigma^p & \text{if } b = 1\,. \end{cases}$$

Claim 1.0.1. $P_{err} = Pr[\hat{b} \neq b] < 1/3$.

Proof. Set $\sigma < n^{-1/p}$, then getting $\epsilon = 0.1$ approximation to $||x * +w||_p^p$ lets Bob learn b with error probability $\delta + o(1) < 1/3$ where δ is the error made in streaming and $o(1)$ is the chance $||x^* + w||_p^p$ deviates by a constant. Thus we will get $P_{err} = Pr[\hat{b} \neq b] < 1/3$. □

Claim 1.0.2. *Information between $\hat{b}$ and b is $\leq \frac{m}{n^{1-2/p}}$.*

Proof. WLOG, we can assume that rows of A are orthonormal. To see this, take SVD of $A = U\Sigma V^T$. Then, if algorithm works using A, it also works using V^T. Given V^T, we can construct Alg' that runs on $V^T x$ as $Alg'(V^T x) = Alg(U\Sigma V^T x)$. Also note that $V^T has \leq m$ rows.

Let v be a row of A, we have

$$I(\langle v, x^* + w\rangle, \langle v, x^*\rangle) = I(\langle v, x^*\rangle + \langle v, w\rangle, \langle v, x^*\rangle)$$

Since w is drawn from a Gaussian distribution, $\langle v, w \rangle$ is additive white Gaussian noise (AWGN). Thus, using Shanon-Hertley, we get

$$I(\langle v, x^* + w \rangle, \langle v, x^* \rangle) \leq \frac{1}{2} \log \left(1 + \frac{E[\langle v, x^* \rangle^2]}{E[\langle v, w \rangle]^2} \right) = \frac{1}{2} \log \left(1 + \frac{1/2n}{\sigma^2} \right) \leq \frac{1}{4n\sigma^2} = \frac{1}{4n^{1-2/p}}$$

since

$$E[\langle v, x^* \rangle^2] = \frac{1}{2} \sum_{i=1}^{n} v_i^2 \cdot \frac{1}{n} = \frac{1}{2n} \qquad\qquad E[\langle v, w \rangle^2] = \sum_{i=1}^{n} v_i^2 \sigma^2 = \sigma^2$$

Note that this only holds for single measurement, we could potentially get more information from multiple observations even if we do not get from one. We will show that information can be bounded additively in this case.

Lemma 1.0.3. *If $a_i = b_i + w_i$ $\forall i \in [n]$ such that each w_i is independent of each other and of b_i, then*

$$I(a, b) \leq \sum_{i=1}^{n} I(a_i, b_i)$$

Proof. We have,

$$\begin{aligned} I(a, b) &= h(a) - h(a|b) = h(a) - h(a - b|b) = h(a) - h(w|b) = h(a) - h(w) \\ &\leq \sum_{i=1}^{n} h(a_i) - h(w_i) = \sum_{i=1}^{n} h(a_i) - h(w_i|b_i) = \sum_{i=1}^{n} h(a_i) - h(a_i - b_i|b_i) \\ &= \sum_{i=1}^{n} h(a_i) - h(a_i|b_i) = \sum_{i=1}^{n} I(a_i, b_i) \end{aligned}$$

Hence, proved. □

Using the above lemma, we get

$$I(A(x^* + w), Ax^*) \leq \frac{m}{4n^{1-2/p}}$$

Since $b \to x^* \to Ax^* \to A(x^* + w) \to \hat{b}$,

$$I(b, \hat{b}) \leq \frac{m}{4n^{1-2/p}}.$$

Hence, proved. □

Using the above claim, we get

$$h(b|\hat{b}) \geq h(b) - \frac{m}{4n^{1-2/p}} = 1 - \frac{m}{4n^{1-2/p}}$$

This shows that even if you know $\hat{b}$, there is sufficient entropy left in b hence still sufficient chance of error. Now, if $m < n^{1-2/p}$, we get $h(b|\hat{b}) \geq 3/4$ and

$$h(b \neq \hat{b}) \geq h(b \neq \hat{b}|\hat{b}) \geq h(b|\hat{b}) \geq 3/4$$

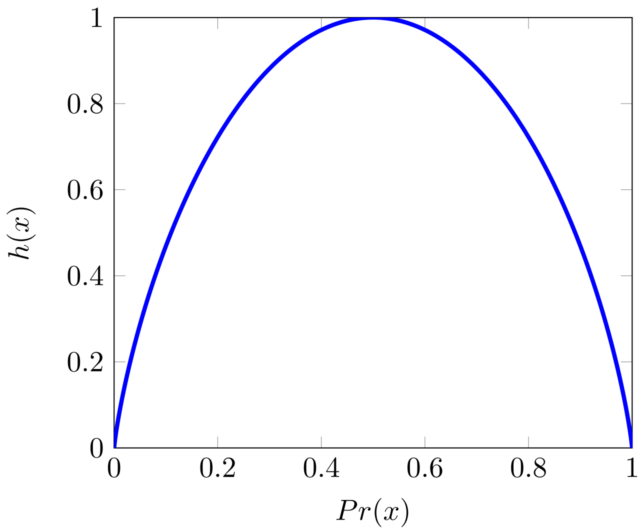

Figure 1: Entropy Curve for bit x

From the entropy curve, it is easy to see that this gives us a contradiction since we have $Pr_{err} = Pr[\hat{b} \neq b] < 1/3$ by the first claim.

2 Packing and Covering Numbers

In this section we discuss packing and covering numbers. We will use results shown here to prove results on compressed sensing in a later lecture.

2.1 Definitions

Definition 2.1.1. *A metric space is an order pair (X, d) where X is a set and $d : X \to \mathbb{R}$ is a distance metric with the following properties for all $x, y \in X$:*

(i) $d(x, y) \geq 0$

(ii) $d(x, y) = 0 \Leftrightarrow x = y$

(iii) $d(x, y) = d(y, x)$

(iv) $\forall z \in X, d(x, y) \leq d(x, z) + d(z, y)$

Definition 2.1.2. *An ϵ-cover of X with respect to d is a collection of points $\{x_1, x_2, \cdots x_n\} \subseteq X$ such that $\forall y \in X \min\limits_{i \in [n]} d(y, x_i) \leq \epsilon$*

Definition 2.1.3. *The covering number $N(\epsilon, X, d)$ is the minimum n such that there exists an ϵ-cover of size n.*

Definition 2.1.4. *The metric entropy of* $\log(N(\epsilon, X, d))$.

We may look at a few examples to understand these definitions.
Example 1: Consider $X = [-1, 1]$ under the distance metric $d(x, y) = |x-y|$. The set $\{0, \pm 2\epsilon, 4\epsilon, \cdots\} \cap [-1, 1]$ is an ϵ-cover. So, we know that $N(\epsilon, X, d) \leq \frac{2}{2\epsilon} + 1$.

Example 2: Consider $X = [-1, 1]^n$ under the distance metric $d(x, y) = \|x - y\|_\infty$. We may view this as covering the box $[-1, 1]^n$ using ϵ^n sized boxes. So, we need at most $(\frac{1}{\epsilon} + 1)^n$ boxes. $N(\epsilon, X, d) \leq (\frac{1}{\epsilon} + 1)^n$.

Definition 2.1.5. *An ϵ-packing of X with respect to d is a collection of points $\{x_1, x_2, \cdots x_n\} \subseteq X$ such that $\min_{i \neq j \in [n]} d(x_j, x_i) \geq \epsilon$*

Definition 2.1.6. *The packing number $M(\epsilon, X, d)$ is the maximum n such that there exists an ϵ-packing of size n.*

In Homework 3, we will prove the following:

Lemma 2.1.7. *$M(2\epsilon, X, d) \leq N(\epsilon, X, d) \leq M(\epsilon, X, d)$*

2.2 Bounds on Covering Number

In this section we study bounds on the covering number of the unit L_q-ball under L_p distance. We wish to bound $N(\epsilon, B_q^d, \| \cdot \|_p)$ where $\|x\|_p = (\sum x_i^p)^{\frac{1}{p}}$ and $B_q^d = \{x \in \mathbb{R}^d \big| \|x\|_p \leq 1\}$.

We know that if $p \geq q$, then we have $\|x\|_p \leq \|x\|_q$. This means $p \geq q \Rightarrow B_q^d \subseteq B_p^d$.

Let us use $\mathsf{Vol}(S)$ to denote the volume of the set S. We note that in $d-$dimensions, $\mathsf{Vol}(aS) = a^d \mathsf{Vol}(S)$

Theorem 2.2.1. $\frac{1}{\epsilon^d} \frac{\mathsf{Vol}(B_q^d)}{\mathsf{Vol}(B_p^d)} \leq N(\epsilon, B_q^d, \| \cdot \|_p) \leq (\frac{2}{\epsilon})^d \frac{\mathsf{Vol}(B_q^d + \frac{\epsilon}{2} B_p^d)}{\mathsf{Vol}(B_p^d)}$

Proof. **Lower Bound:**
Let $x_1, \cdots, x_n$ be an ϵ-cover of B_q^d under the L_p norm.
Then,

$$B_q^d \subseteq \bigcup_{i \in [n]} (x_i + \epsilon(B_p^d))$$

So, we may bound the volume:

$$\begin{aligned} \mathsf{Vol}(B_q^d) &\leq n \cdot \mathsf{Vol}(\epsilon B_p^d) \\ &= n \cdot \epsilon^d \cdot \mathsf{Vol}(B_p^d) \end{aligned}$$

So, $n = N(\epsilon, B_q^d, \| \cdot \|_p) \geq \frac{1}{\epsilon^d} \frac{\mathsf{Vol}(B_q^d)}{\mathsf{Vol}(B_p^d)}$

Upper Bound:

Let $x_1, \cdots, x_m$ be an ϵ-packing of B_q^d.

This means that the balls $x_i + \frac{\epsilon}{2} B_p^d$ are disjoint (except at the surface). We also know that all of these balls are contained in $B_q^d + \frac{\epsilon}{2} B_p^d$ since some of these balls might be on the surface of B_q^d.

So,

$$\begin{aligned} \mathsf{Vol}(B_q^d + \frac{\epsilon}{2} B_p^d) &\geq m \cdot \mathsf{Vol}(\frac{\epsilon}{2} B_p^d) \\ &= m \cdot (\frac{\epsilon}{2})^d \cdot \mathsf{Vol}(B_p^d) \end{aligned}$$

So, we conclude that $m = M(\epsilon, B_q^d, \| \cdot \|_p) \leq (\frac{2}{\epsilon})^d \cdot \frac{\mathsf{Vol}(B_q^d + \frac{\epsilon}{2} B_p^d)}{\mathsf{Vol}(B_p^d)}$

Using Lemma 2.1.7, we may conclude that $N(\epsilon, B_q^d, \| \cdot \|_p) \leq (\frac{2}{\epsilon})^d \cdot \frac{\mathsf{Vol}(B_q^d + \frac{\epsilon}{2} B_p^d)}{\mathsf{Vol}(B_p^d)}$

□

Let us examine this Theorem under the values $p = 2$ and $q = 1$.

$\mathsf{Vol}(B_1^d) = \frac{2^d}{d!}$ and $\mathsf{Vol}(B_2^d) = \frac{\pi^{\frac{d}{2}}}{(d/2)!}$ when d is even.

The bound on $N = N(\epsilon, B_1^d, \| \cdot \|_2)$ which we get by applying this theorem can be written in terms of metric entropy as:

$$d \log(1/\epsilon) - \frac{d}{2} \log(d) \leq \log(N) \leq d \log(\frac{2}{\epsilon})$$

Observe that this gives us a tight bound when $\epsilon << \frac{1}{d}$. However, when $\epsilon >> \frac{1}{\sqrt{d}}$, the bound is trivial.

Intuitively, we may explain this as follows: The L_1 ball in $d-$dimensions should be viewed as being "spikey". The ϵB_2^d balls, which we use to cover the tips of these spikes incur a huge loss in volume i.e a large volume which we account for in the bound is outside the volume which we wish to cover.

We will use a different method in order to prove that $\log(N) \leq \frac{1}{\epsilon^2} \log(d)$ in the next lecture.